

ABSTRAK

Banyaknya jumlah kematian dini akibat jantung, salah satunya terjadi karena terlambatnya penanganan pasien dengan risiko tinggi. Kematian dini terjadi karena pasien tersebut tidak masuk dalam kelompok prioritas, yang berdampak pada terlambatnya proses penanganan. Dari hal tersebut bisa dikatakan bahwa, dengan mengklasifikasi pasien risiko tinggi ke dalam kelompok prioritas, dapat mencegah meningkatnya jumlah kematian dini akibat gagal jantung. Sehingga dilakukan penelitian untuk mengklasifikasikan ketahanan hidup pasien gagal jantung, dengan menerapkan *balancing* data dan seleksi atribut, yang mana dapat membantu tenaga medis untuk menentukan prioritas penanganan yang akan dilakukan.

Penelitian dilakukan ke dalam beberapa tahap. Pada tahap pertama, *missing value* atau inkonsisten yang terdapat dalam data akan diperbaiki (diubah/dihapus) dahulu sebelum *balancing*. Proses *balancing* menggunakan metode ADASYN, yang digunakan untuk melakukan *oversampling* pada kelas minoritas, karena jumlah data pasien yang meninggal lebih sedikit dibandingkan jumlah data pasien yang bertahan hidup. Data yang sudah *balance* kemudian dinormalisasi dengan mengubah *range* data setiap atribut ke dalam *range* yang sama. Data yang telah dinormalisasi, kemudian diseleksi untuk mendapatkan atribut-atribut yang paling relevan. Adapun metode seleksi menggunakan *one-way* ANOVA dengan memberi peringkat pada setiap atribut berdasarkan jumlah nilai varian rata-ratanya (F_0), lalu dipilih nilai atribut yang memenuhi nilai F dalam tabel distribusi (F_t) di mana $F_t = 7,96765$. Data hasil *balancing* dan seleksi ini kemudian digunakan dalam perhitungan akurasi dan pengujian data menggunakan metode *Random Forest*.

Klasifikasi yang dilakukan terhadap data *balancing* dengan atribut *time*, *serum creatinine*, dan *ejection fraction* menggunakan *7-fold cross validation*, menghasilkan rata-rata akurasi paling tinggi yaitu 89,11%. Sedangkan untuk *5-fold cross validation* dan *3-fold cross validation* masing-masing sebesar 87,46% dan 87,50%.

Kata Kunci : Gagal Jantung, *Random Forest*, ADASYN, *One-way* ANOVA, Klasifikasi.

ABSTRACT

The large number of premature deaths due to heart disease, one of which occurs due to the delay in handling high-risk patients. Early death occurred because the patient was not included in the priority group, which resulted in delays in the treatment process. From this it can be said that, by classifying high-risk patients into priority groups, it can prevent an increase in the number of premature deaths due to heart failure. So that a study was conducted to classify the survival of heart failure patients, by applying data balancing and attribute selection, which can help medical personnel to determine the priority of treatment to be carried out.

The research was carried out in several stages. In the first stage, missing or inconsistent values contained in the data will be corrected (changed/deleted) before being balanced. The balancing process uses the ADASYN method, which is used to oversampling the minority class, because the number of data on patients who died is less than the number of data on patients who survive. The balanced data is then normalized by changing the data range of each attribute into the same range. The normalized data is then selected to obtain the most relevant attributes. The selection method uses one-way ANOVA by ranking each attribute based on the number of average variance values (F_o), then selecting the attribute values that meet the F value in the distribution table (F_t) where $F_t = 7.96765$. The data from the balancing and selection results are then used in calculating the accuracy and testing the data using the Random Forest method.

Classification carried out on the data after balancing with attribute time, serum creatinine, and ejection fraction using 7-fold cross validation, resulted in the highest average accuracy of 89.11%. Meanwhile, for 5-fold cross validation and 3-fold cross validation, they are 87.46% and 87.50%, respectively.

Keywords: Heart Failure, *Random Forest*, ADASYN, *One-way* ANOVA, Classification.